

Обучение машинного перевода без параллельных текстов*

Ярошенко А. М. Бахтеев¹ О. Ю. Стрижов² В. В.

¹Московский физико-технический институт

²Вычислительный центр им. А. А. Дородницына ФИЦ ИУ РАН

В данной работе исследуется задача машинного перевода между двумя языками. Для решения часто используются параллельные предложения, то есть совпадающие по смыслу фразы на двух языках. В работе рассматривается альтернативная модель, не требующая большого количества параллельных предложений. Она использует нейронную сеть типа Seq2Seq, имеющую скрытое пространство. [Тут добавится что-то от меня]. Для проверки качества модели проводится вычислительный эксперимент по переводу предложений между близкими языками, такими как русский и украинский.

Ключевые слова: *нейронные сети, машинный перевод, автокодировщики.*

Введение

В зависимости от специфики пары языков выделяют несколько подходов к машинному переводу. При наличии достаточного числа параллельных предложений (порядка миллиона) использование глубоких нейронных сетей привело к получению хороших результатов [1],[2].

Но для многих пар языков нет достаточной базы примеров. Одним из подходов на основе параллельных предложений является пополнение обучающей выборки переводами с предыдущих итераций работы нейронной сети [3].

Ниже представлено решение задачи машинного перевода при отсутствии достаточного количества параллельных предложений [4], [5], [6]. В модели используются 2 типа автокодировщиков: рекуррентные нейронные сети LSTM ([7], [8]), которые реализуют перевод слов в скрытое пространство, и сеть-дискриминатор, определяющая по векторному представлению язык исходного предложения. Сети LSTM оптимизируются так, чтобы представление одного и того же предложения на разных языках совпадало в скрытом пространстве, то есть, чтобы дискриминатору было сложнее определить язык, к которому относится вектор. Обучение состоит из двух фаз. На первой оптимизируется работа дискриминатора: предложение кодируется с добавлением шума ([9]) и подаётся на вход и происходит перераспределение параметров. На второй стадии происходит перераспределение параметров уже у сетей-кодировщиков. После проведения этих шагов вычисляется значение функции потерь.

Такой подход был использован в [10] для пары языков французский-русский. В данной работе будет проведен схожий эксперимент для перевода с русского на украинский. Качество переводчика в работе оценивается с помощью метрики BLEU [11].

Постановка задачи

В данной задаче в качестве обучающей выборки используются несопоставленные друг другу предложения на обоих языках $D^{src} = [s_1^{src}, \dots, s_{m_{src}}^{src}]$, $D^{tgt} = [s_1^{tgt}, \dots, s_{m_{tgt}}^{tgt}]$, по которым

нужно предоставить перевод на другой язык. Также предоставлен блок параллельных предложений для проверки качества перевода.

Предлагается решение в виде модели из двух рекуррентных нейронных сетей для реализации декодера, дискриминатора и энкодера. Для нулевого приближения используется пословный перевод [12].

Ошибка модели на валидационной выборке складывается из трех составляющих: доли в целом неправильно переведенных предложений, доли ошибочно переведенных слов и ассигасы, параметры которой будут подобраны в ходе эксперимента (я плохо поняла, мы выберем одну из них? если да, то я за первую).

Так как у нас нет достаточно большого корпуса из параллельных предложений, мы будем использовать следующую схему построения модели. Первая нейронная сеть энкодер будет переводить исходное предложение в скрытое пространство, где дискриминатор будет по вектору определять, какому языку он принадлежал и соответственно использовать декодер, соответствующий другому языку. Идея этого решения в том, чтобы приблизить друг к другу пространства, соответствующие разным исходным языкам.

Для реализации этого метода определим функционалы, которые будут минимизироваться. Во-первых необходимо зашумить исходные предложения, чтобы модель не обучилась возвращать в конце цикла исходные данные. Пусть $\sigma(x)$ - результат наложения шума на слово x . На этом шаге оптимизации будет минимизироваться следующая функция:

$$L_{AE} = ||d(e(\sigma(x))) - x||^2$$

Далее на этапе использования пословного перевода функция потерь будет иметь вид:

$$L_{TR} = ||d(e(\hat{g}(e(x)))) - x||^2$$

И последний этап - оптимизация дискриминатора, чтобы он различал представления векторов разных языков в скрытом пространстве:

$$L_{ADV} = \log p(\text{lang} = \text{src} | \text{Encoder}(x)) + \log p(\text{lang} = \text{tgt} | \text{Encoder}(y))$$

Таким образом, имеем задачу оптимизации:

$$L = a * L_{AE} + b * L_{TR} + c * L_{ADV} \rightarrow \min$$

где a, b, c калибруемые гиперпараметры.

Литература

- [1] Bilingual word embeddings for phrase-based machine translation / Will Y Zou, Richard Socher, Daniel Cer [и др.] // Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013. С. 1393–1398.
- [2] On the properties of neural machine translation: Encoder-decoder approaches / Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau [и др.] // arXiv preprint arXiv:1409.1259. 2014.
- [3] Bertoldi Nicola, Federico Marcello. Domain adaptation for statistical machine translation with monolingual resources // Proceedings of the fourth workshop on statistical machine translation / Association for Computational Linguistics. 2009. С. 182–189.
- [4] Google's neural machine translation system: Bridging the gap between human and machine translation / Yonghui Wu, Mike Schuster, Zhifeng Chen [и др.] // arXiv preprint arXiv:1609.08144. 2016.

- [5] Sutskever Ilya, Vinyals Oriol, Le Quoc V. Sequence to sequence learning with neural networks // Advances in neural information processing systems. 2014. С. 3104–3112.
- [6] Bahdanau Dzmitry, Cho Kyunghyun, Bengio Yoshua. Neural machine translation by jointly learning to align and translate // arXiv preprint arXiv:1409.0473. 2014.
- [7] Gers Felix A, Schmidhuber Jürgen, Cummins Fred. Learning to forget: Continual prediction with LSTM. 1999.
- [8] Graves Alex, Schmidhuber Jurgen. Framewise phoneme classification with bidirectional LSTM and other neural network architectures // Neural Networks. 2005. Т. 18, № 5-6. С. 602–610.
- [9] Kim Yunsu, Ney Jiahui Geng Hermann. Improving Unsupervised Word-by-Word Translation with Language Model and Denoising Autoencoder. 2018.
- [10] Lample Guillaume, Denoyer Ludovic, Ranzato Marc'Aurelio. Unsupervised machine translation using monolingual corpora only // arXiv preprint arXiv:1711.00043. 2017.
- [11] BLEU: a method for automatic evaluation of machine translation / Kishore Papineni, Salim Roukos, Todd Ward [и др.] // Proceedings of the 40th annual meeting on association for computational linguistics / Association for Computational Linguistics. 2002. С. 311–318.
- [12] Word translation without parallel data / Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato [и др.] // arXiv preprint arXiv:1710.04087. 2017.